

2000

OLAP

2001

OLAP

論文 碩士學位 論文 提出

2001 年 01 月

梨花女子大學校 科學技術大學院

學科 金 帛 宣

碩 士 學 位 論 文 認 准

指導教授

審查委員

_____	_____
_____	_____
_____	_____

梨 花 女 子 大 學 校 科 學 技 術 大 學 院

	vi
I	1
1.1	1
1.2	3
II	5
2.1	5
2.2	7
2.2.1	8
2.2.2	9
2.2.3	9
2.2.4	10
2.2.5	11
2.2.6	12
2.3	13
2.3.1	13
2.3.2 WEBMINER	14
2.3.3 WUM (Web Utilization Miner)	14
2.3.4 Web Log Miner	16
2.3.5 WebSIFT (Web Site Information Filter)	16
III .	OLAP	18
3.1 W3C	18
3.2	20
3.2.1	22

3.2.2	23
3.2.3	23
3.3	25
3.4 OLAP	29
IV	33
4.1	33
4.2	34
4.3	35
4.3.1	36
4.3.2	37
4.3.3	38
4.3.4	39
4.3.5 OLAP	44
4.3.6	46
4.4	47
4.4.1	48
4.4.2	51
V	55
	58
	60

2-1	7
2-2	:	8
2-3	:	11
2-4	:	12
2-5 WUM	15
3-1	21
3-2	22
3-3	23
3-4	24
3-5	25
3-6	26
3-7	27
3-8 IP	28
3-9	29
3-10	31
3-11	31
4-1	35
4-2	37
4-3	37
4-4	38
4-5	40
4-6	40
4-7	41
4-8	41
4-9	42

4-10	OLAP	43
4-11		43
4-12		44
4-13	DTS Browse : DTS	45
4-14		1	47
4-15		2	48
4-16		3	49
4-17	K-	50
4-18		51
4-19		51

3-1 W3C	17
3-2	18
4-1	28
4-2	36

(Web Usage Mining)

(Data Mining)

가

IIS

3

(On-Line Analytical Processing)

OLAP

(On-

Line Analytical Processing Mining Model)

OLAP

5

Microsoft SQL Server 2000

Data Transformation Services Designer

Analysis Services Manager

I.

1.1

(Web Mining)

(Data Mining)

가

(Web Usage Mining)

[3,9].

가

가 ,

(Association Rule),

(Clustering),

(Classification),

(Sequential

Pattern)

[5,10].

(Business Intelligent) , 가
(Usage Characterization), (Personalization) .

가

(Extraction), (Transformation), (Load) .

, OLAP ,

가

CRM(Customer Relationship Management)

, OLAP

80%

가

1.2

OLAP

가

Microsoft Data Transformation Services Designer

Microsoft SQL2000 Analysis Management

Services

.

. 2

. 3

OLAP

4

.

5

.

II.

2.1

(Web Contents Mining),

(Web Structure Mining)

[3, 4, 5, 8, 9].

(Server-level collection),

(Client-level collection),

(Proxy-level collection)

가

가

(Cached page view)

, POST

HTTP

가

(User Privacy)

가

CGI

가

가

(Java applet, Java script...)

가

2-1

3

[2, 3].

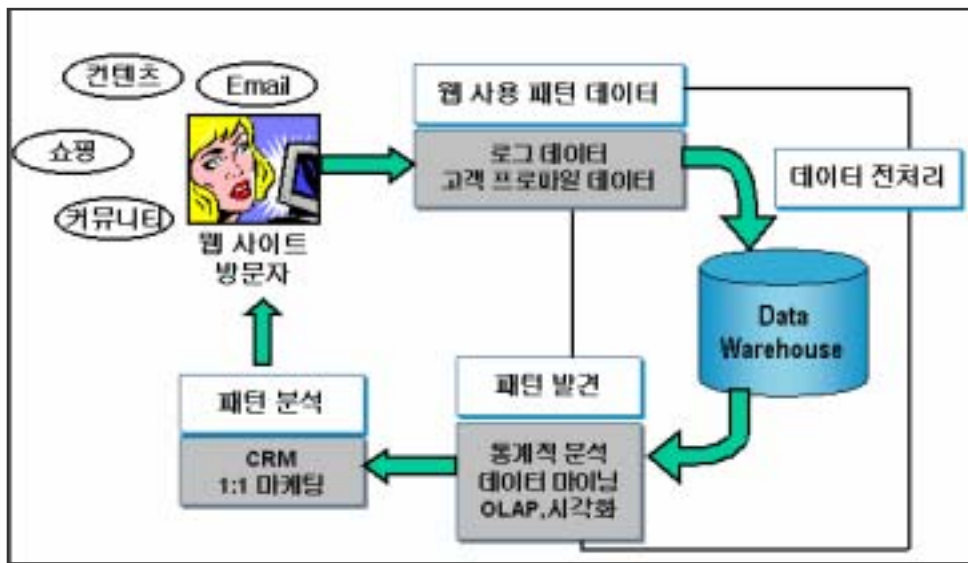
가

OLAP

(Business Intelligent),

(Usage Characterization)

[3, 4, 5].



2-1.

가

가

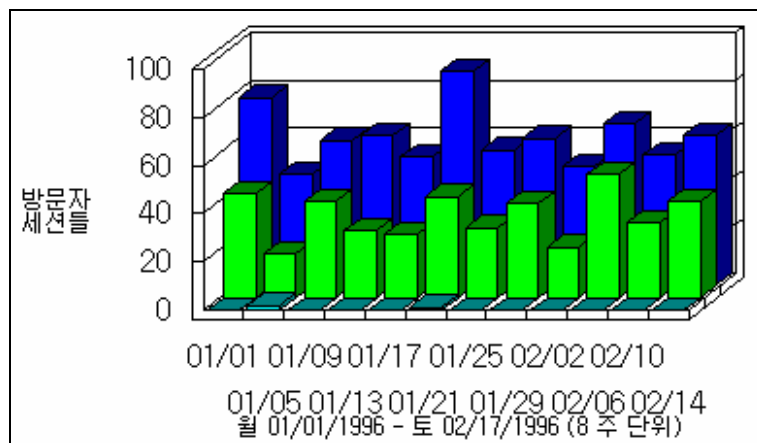
2.2

가 [3, 20].

2.2.1

가 , (viewing time), (navigational path) 가 (, , ..)

URI



2-2.

:

2.2.2 (Association Rule)

가 .
 (threshold) (support value)

. , Apriori

(latency)

가 (document fetching)

가 .

- | | |
|--|-----|
| 1) /company/products/product1.html | 40% |
| /company/products/product2.html | |
| 2) /company/announcements/special-offer.html | 30% |
| /company/products/product1 | |

(prune) .

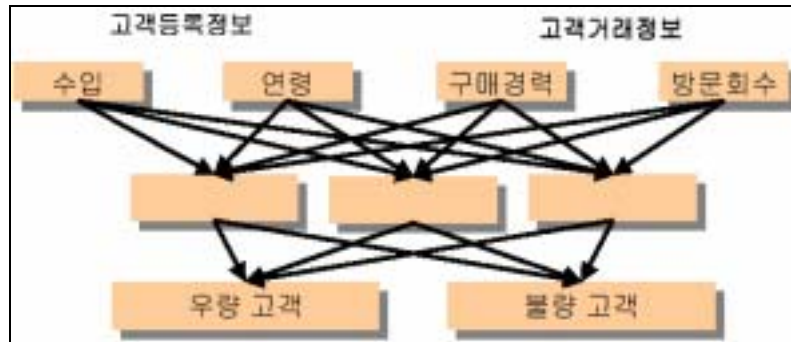
2.2.3 (Clustering)

가 , 가
 가
 (market segmentation) ,
 (personalized web content)
 가
 (web assistance provider)
 html
 가

- 1) /company/products/ product1.html
- 2) /company/products/product2 50% 20
 25

2.2.4 (Classification)

(mapping)
 가
 (naïve Bayesian), K-
 (K-nearest neighbor) (supervised inductive
 learning algorithm) , 2-3



2-3. :

2.2.5 (Sequential Pattern)

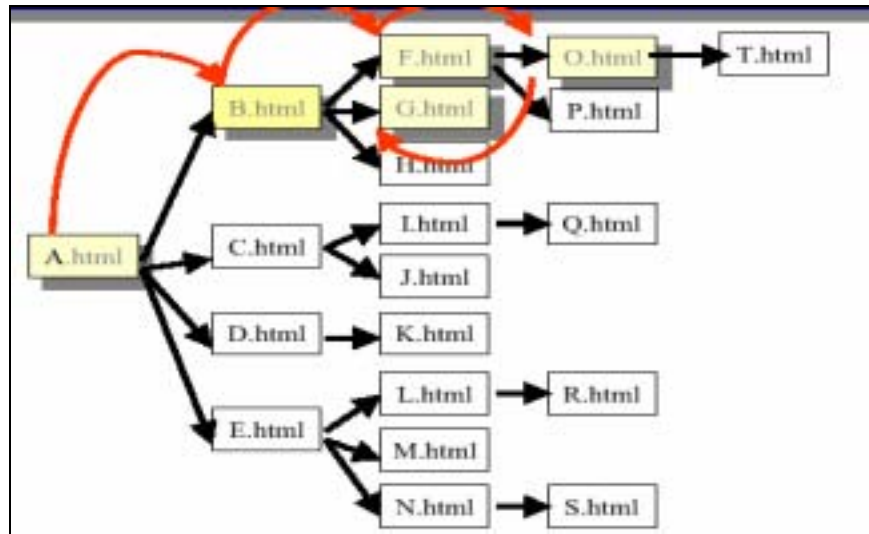
1) /company/products/ 30%

W

2) /company/products/product1.html

60% 15 /company/products/product4.html

(changing point detection),



2-4. :

2.2.6 (Dependency Modeling)

Bayesian Belief Network

Hidden Markov Model,

가

2.3 (Web Usage Mining System)

30

WEBMINER, WebLogMiner, WUM, WebSIFT

가 .

가 가

2.3.1

WebTrends, WebLog, SpeedTracer, analog, WUAGE,

WWWSTAT, SEBTREND, FASTSTATS, WEBALIZER [3, 16, 17, 18].

, , , , ,

가 .

가 .

(gif, jpg, avi, ...)

가

가 가 .

가 가
가

. ,
, 가 .

2.3.2 WEBMINER

(University of Depaul)

, ,

(access log),

(referrer log), (agent log) , html

가 .

, ,

(path completion) ,

,

[2].

2.3.3 WUM (Web Utilization Miner)

(Humboldt University)

MINT .

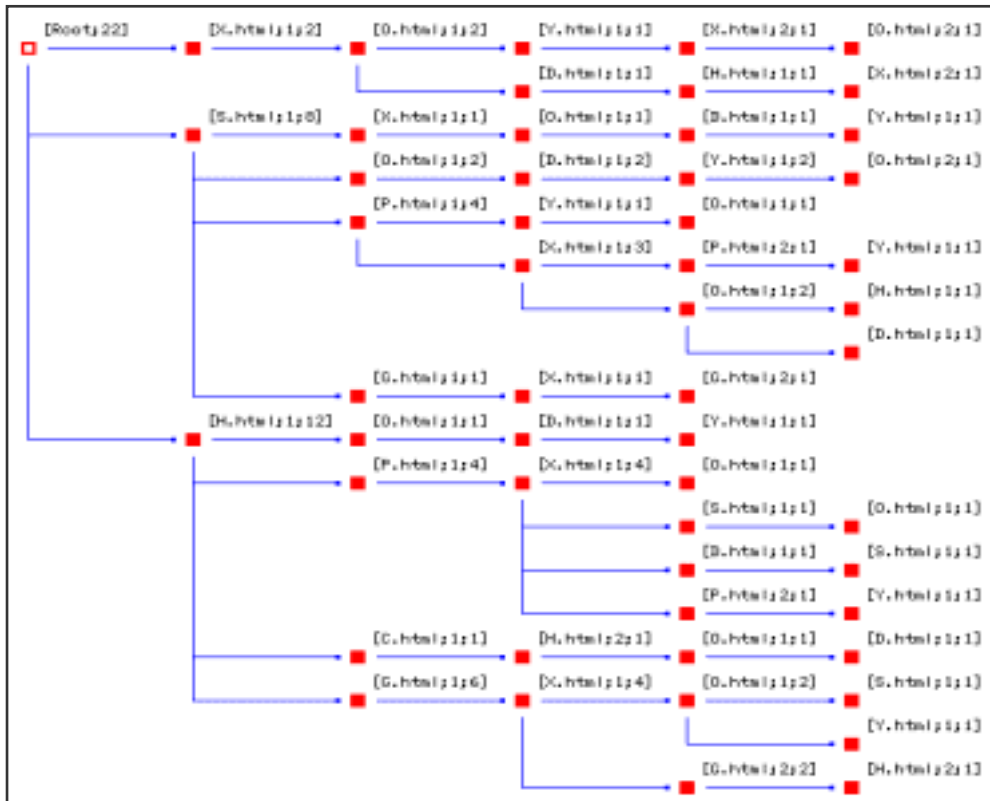
(

)
,
.

WUM

(prefix) 가

2-5 [18].



2-5. WUM

2.3.4 Web Log Miner

(Simon Fraser University)

Virtual-U research laboratory TeleLearning-NCE Intelligent DBSystem research
 laboratory NCE-IRIS

. WebLogMiner 4 가 . 1

2

. 3 (On-Line

Analytical Processing) 4

.

,

[1].

2.3.5 WebSIFT (Web Site Information Filter)

University of Minnesota (Web Usage Mining),

(Web Content Mining), (Web Structure Mining)

.

(quantitative model)

[4].

,

.

(access log),

(referrer log), (agent log) NCSA

.

,

(Information Filter)

III.

OLAP

IIS(Internet Information Services)

W3C

OLAP

OLAP

OLAP

3.1 W3C

W3C

W3

HTTP

가

CR

‘-‘

. W3C

(Directive)

3-1

3-1. W3C

date	
time	
c-ip	IP
c-username	
s-sitename	
s-computername	
s-ip	IP
cs-method	가 (:GET)
cs-uri-stem	(:HTML , CGI ,)
cs-uri-query	가 .
sc-status	HTTP
sc-win32-status	Windows
cs-bytes	가
sc-bytes	가
s-port	가
time-taken	
cs-protocol	가 (HTTP, FTP)
cs(User-Agent)	
cs(Cookie)	.
cs(Referer)	

IIS 가 Microsoft Internet Information Services

(IIS Log Format), NCSA(National Center for Supercomputing Applications)

(Common Log Format : CLF), W3C (Extended Log Format), ODBC Logging

[21]. CERN CLF 가 가 ,

W3C ,

가

(Back-Tracking),

(Reload)

가

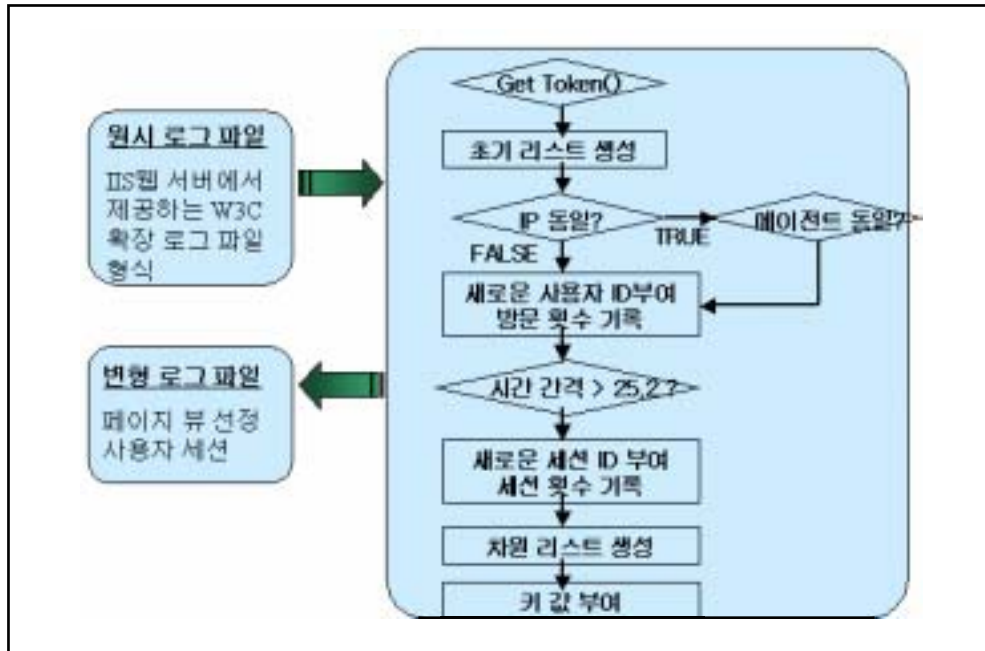
W3C

3.2

OLAP

가

3-1



3-1.

3-2

WCA(Web Characterization

Activity)

3-2.

	가
	가 가
	(: KB)

3.2.1 (Data Cleansing Module)

. html
 .
 가 cs_uri_stem
 html ,
 가 gif, jpeg, jpeg, map avi, cgi / /
 . 3-2
 1/10 ~ 1/40 가
 가 .

```

BOOL IsaPage(char *str)
{
  cs_uri_stem
  if (cs_uri_stem 가 ".htm" )) found = TRUE;
    else if (cs_uri_stem 가 ".html")) found = TRUE;
    else if (cs_uri_stem 가 ".asp")) found = TRUE;
    else if (cs_uri_stem 가 ".shtml")) found = TRUE;
    else if (cs_uri_stem 가 ".phtml")) found = TRUE;
    else if (cs_uri_stem 가 ".php3")) found = TRUE;
    else if (*(temp + strlen(temp) - 1) == '/') found = TRUE;
    else if (strstr(temp, ".") == NULL) found = TRUE;
}
  
```

3.2.2 (User Identification Module)

가 IP
 c_ip cs_agent
 c_ip cs_agent
 , cs_username cs_uri_query
 ID
 ID (: Guest00001). ID http
 3-3

```

void WriteUserIDToMainList(void)
{
    while(c_ip cs_user_agent ) {
        if ( ID 가 )
            ID cs_username ;
        else if ( ID 가 )
            InsertGuest( ID );
    }
}
    
```

3-3.

3.2.3 (Session Identification Module)

Time date

25.2 (1512) ID . (: Session00001).

25.2

[2].

3-4

3-5

```

void WriteSessionToMainList(void)
{
    while(
        if (strcmp(
            if (strcmp(
                time1 =
                time2 =
                if ((time2 - time1) > 1512) session++;
            } else
                session++;
        } else (
            _itoa(
            InsertSession(
    }

```

3-4.

```

2000-11-01 07:21:03 128.134.121.184 guest0001 W3SVC8 S4TS_SUB2 211.118.48.11
GET /SportsCenter/diet/0024.asp - 200 0 14862 422 578 80 HTTP/1.0
Mozilla/2.0+(compatible;+MSIE+3.0;+Windows+98;+DigExt) ASPSESSIONIDQGGQ
http://www.s4ts.Session000 Mozilla/2.0+(compatible MSIE+3.0 - -
/SportsCenter/diet/0001.asp /SportsCenter/diet/diet.asp

2000-11-01 07:20:47 128.134.121.184 guest0001 W3SVC8 S4TS_SUB2 211.118.48.11
GET /SportsCenter/diet/0015.asp - 200 0 13070 422 203 80 HTTP/1.0
Mozilla/2.0+(compatible;+MSIE+3.0;+Windows+98;+DigExt) ASPSESSIONIDQGGQ
http://www.s4ts.Session000 Mozilla/2.0+(compatible MSIE+3. - - /SportsCenter/diet/0001.asp
/SportsCenter/diet/diet.asp

2000-11-02 16:20:30 128.134.130.11 guest0002 W3SVC8 S4TS_SUB2 211.118.48.11 GET
/inde x.asp - 200 0 40161 86 1844 80 HTTP/1.0 lhWeb/c/lsHTTP+--+hiongun@kt.co. - -
Session000 lhWeb/c/lsHTTP+--+hiongun@kt.co. - - - /inde x.asp /inde x.asp

2000-11-01 16:36:05 128.134.130.11 guest0002 W3SVC8 S4TS_SUB2 211.118.48.11 GET
/inde x.asp - 200 0 39195 86 1719 80 HTTP/1.0 lhWeb/c/lsHTTP+--+hiongun@kt.co. - -
Session001 lhWeb/c/lsHTTP+--+hiongun@kt.co. - - - /inde x.asp /inde x.asp
    
```

3-5.

3.3

	cs_uri_stem	‘/’	3-6
	date/time		
3-7	ip	c_ip	ip
	가		-3.8
	(key)		
3-9			

가

```

typedef struct SiteList{           //
    char *path;                     //
    char *site_key;                 //
    PathListS *child;              //
    struct SiteList *next;
} SiteListS;

void MakeSiteDimension(void) //
{
    SiteListS *temp,    PathListS *pathNode,    max_path_no = 0;
    SiteRoot = NULL,   mCurrent = MainRoot;

    if (mCurrent){
        InsertSiteList(temp); // cs_uri_stem
        mCurrent = mCurrent->next;
        while (mCurrent){
            sCurrent = SiteRoot;
            while ( 가 ){
                sCurrent = sCurrent->next;
                if ( 가 ){
                    CreateSiteList(temp);
                    temp->path = mCurrent->cs_uri_stem);
                    // ' GetToken()
                    // child 가 child
                    temp->child = MakePathList(temp->path.);
                    InsertSiteList(temp); }
                } mCurrent = mCurrent->next; /* else*/
            }
        }
        SortSiteList();
        InsertKeyS(keyID);
        InsertPathList(pathNode, "-"); // child
    }

void SaveSiteDimension(void) //

```

```

typedef struct TimeList{ //
    char *the_month; //
    char *the_day; //
    char *date; //
    char *time; //
    char *time_key; //
    struct TimeList *next;
} TimeListS;
void MakeTimeDimension(void) //
{
    int current_month; // date
    int current_day; // date
    TimeListS *temp;
    MakeDateList(); //
    SortDateList(); //
    TimeRoot = NULL;
    dCurrent = DateRoot;
    while (dCurrent){
        current_month = GetMonth(dCurrent->date);//
        current_day = GetDay(dCurrent->date); //
        // 24
        for (i = 0; i < 24; i++){
            temp = CreateTimeList();
            temp->the_month = month[current_month];
            temp->the_day = day[current_day];
            temp->date = dCurrent->date;
            temp->time = time[i];
            InsertTimeList(temp);
        }
        dCurrent = dCurrent->next;
    }
    InsertKeyT(keyID);
}
void SaveTimeDimension(void) //

```

```

typedef struct IPList{           // IP
    char *ip;                    // ip
    char *ip_count;              //      ip
    char *rank;                  //
    char *ip_key;                //      IP
    struct IPList *next;
} IPListS;
void MakeIPDimension(void)
{
    IPListS *temp, IPRoot = NULL, mCurrent = MainRoot;
    if (mCurrent){
        InsertIPList(temp);
        mCurrent = mCurrent->next;
        while (mCurrent){           //      IP      ip_count      가
            if (      ip      )      count++;
            else if (      ip      ){
                iCurrent = iCurrent->next;
                temp = CreateIPList();
                temp->ip = strdup(mCurrent->c_ip);
                temp->ip_count = ItoA(1);
                InsertIPList(temp);
            }
            mCurrent = mCurrent->next;
        }
    }
    SortIPList();
    for (i = 0; iCurrent && i < 50; i++){
        iCurrent->rank = "Frequent";
        InsertKeyI(keyID);
        iCurrent = iCurrent->next; }
    while (iCurrent){
        iCurrent->rank = "Rare";
        InsertKeyI(keyID);
        iCurrent = iCurrent->next; }
    }
void SaveIPDimension(void)      // IP

```

3-8. IP

```

2000-11-02 08:37:20 128.110.29.180 guest00000 W3SVC8 S4TS_SUB2 211.118.48.11 GET
/sportsscience/sportsscience_index.asp - 200 0 17691 431 1640 80 HTTP/1.1 - http://science.s
Session000 Mozilla/4.0+(compatible MSIE+5.01 Windows+NT+5.0 - -
/SportsScience/psychology/Psychology_3.asp /sportsscience/sportsscience_index.asp 1 0 0
C000000041 S000000591 T000000128 I000004784

2000-11-01 07:21:03 128.134.121.184 guest00001 W3SVC8 S4TS_SUB2 211.118.48.11
GET /SportsCenter/diet/0024.asp - 200 0 14862 422 578 80 HTTP/1.0
ASPSESSIONIDQGGQ http://www.s4ts. Session000 Mozilla/2.0+(compatible MSIE+3.0
Windows+98 - - /SportsCenter/diet/0001.asp /SportsCenter/diet/diet.asp 1 0 0 C000001487
S000000143 T000000103 I000004302

2000-11-01 07:20:47 128.134.121.184 guest00001 W3SVC8 S4TS_SUB2 211.118.48.11 GET
/SportsCenter/diet/0015.asp - 200 0 13070 422 203 80 HTTP/1.0 ASPSESSIONIDQGGQ
http://www.s4ts. Session000 Mozilla/2.0+(compatible MSIE+3.0 Windows+98 - -
/SportsCenter/diet/0001.asp /SportsCenter/diet/diet.asp 1 0 0 C000001487 S000000134
T000000103 I000004302
    
```

3-9.

3.4 OLAP

OLAP

가 ,

가 . OLAP

(member

properties) 가

가

가 . 가
 가
 Microsoft Analysis Services 2000 가

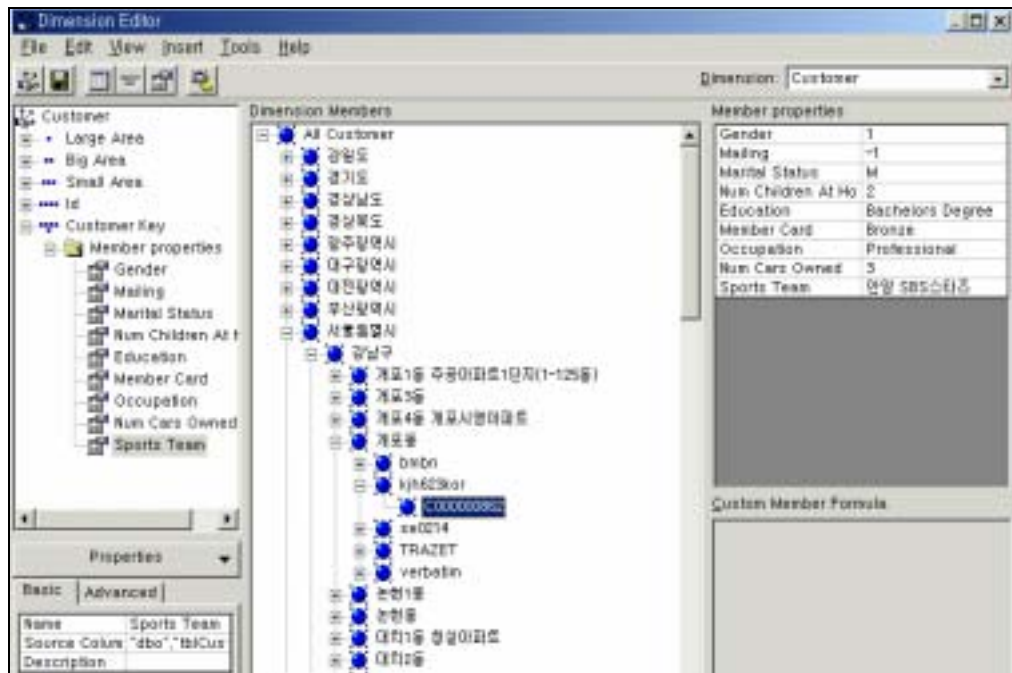
.
 ,
 , IP , 4
 , 3-10 3 가
 가 .

IP IP IP

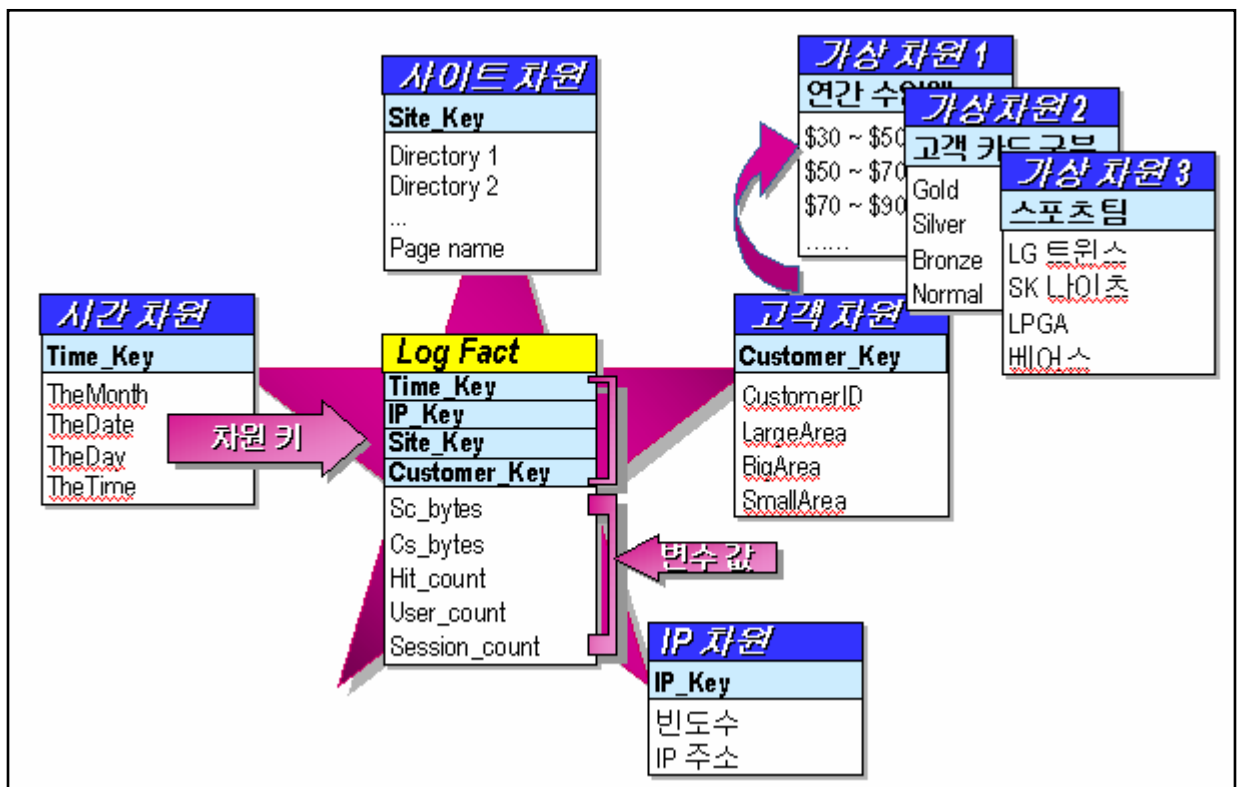
(TheMonth), (TheDay), (TheDate), (TheTime) 4

(SmallArea), ID (CustomerID) 4
 (LargeArea), (BigArea),

가 , 가
 (Gold), (Silver), (Bronze), (Normal)



3-10.



3-11

OLAP

(Case)

(weighting)

.

(Gender),

(Mailing),

(Marriage),

(YearlyIncome),

(TotalChildrenAtHome),

(Education),

(MemberCard),

(CarsOwned),

(SportsTeam)

,

,

.

IV.

3

OLAP

OLAP

4

4.1

4-1

[4.1]

	Windows 2000 Server
	Microsoft SQL 2000 server
	Internet Information Server 5(IIS5)
	Microsoft Visual C/C++ 6.0 Microsoft Visual Basic 6.0 Microsoft SQL, OLE DB API, DTS Package Designer

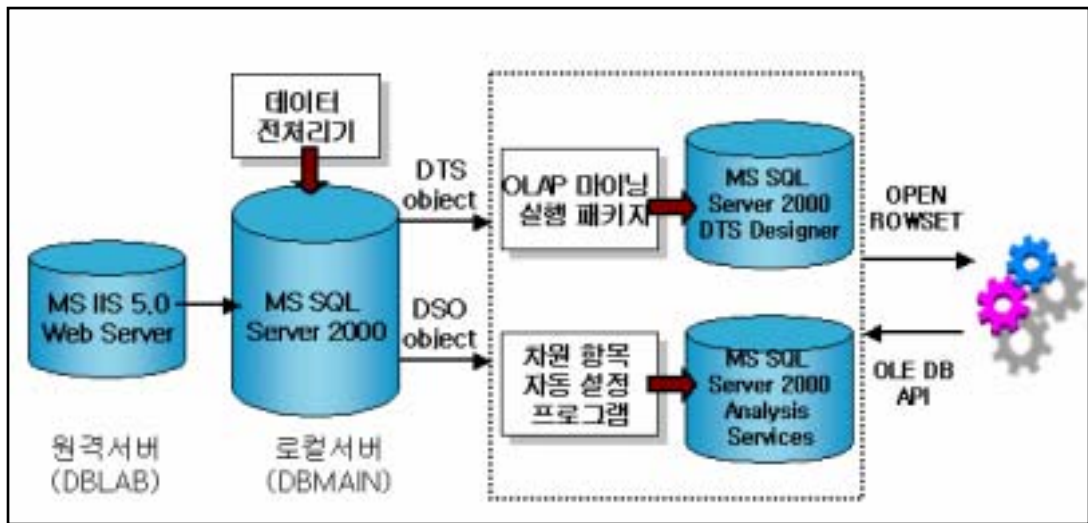
IIS

Microsoft Visual C/C++ ,
 Microsoft SQL 2000 Server Analysis Services Manager Data Transformation
 Services Designer . Analysis
 Service API OLE DB API 가 . Decision Support Object
 OLAP ,
 Microsoft Visual Basic ActiveX script Microsoft SQL [13, 14].

4.2

. DTS OLAP
 DTS Designer . DSO Analysis Services
 , OLE DB API OLE DB for Data
 Mining OPENROWSET

4-1 .



4-1.

4.3

OLAP

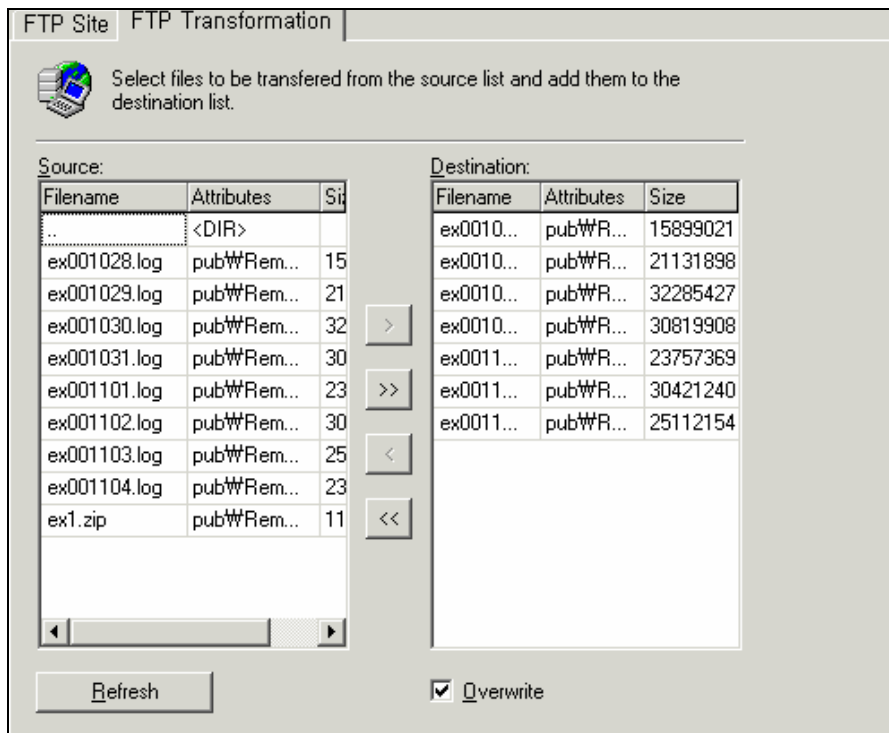
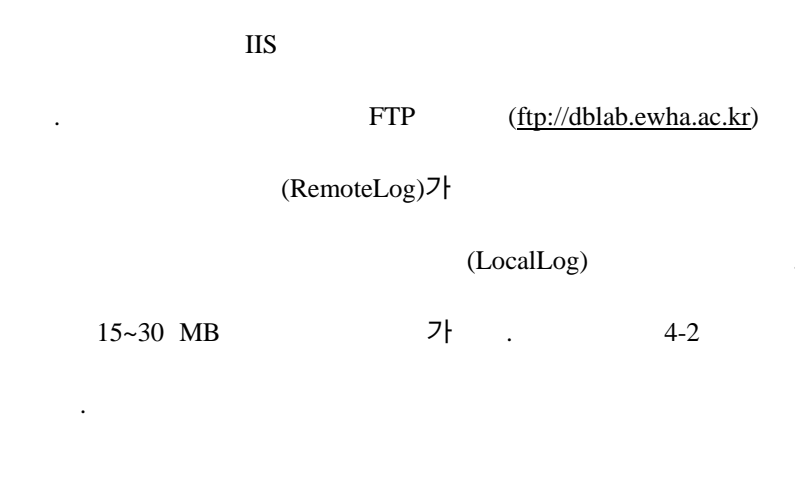
, OLAP

5

가

가

4.3.1



4-2.

4.3.2

. W3C

가 .

4-3 .

(Connection 1)

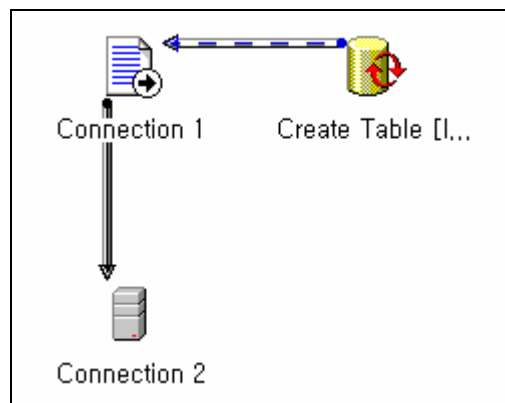
(Connection 2)

4-4 Data Transformation Services Object

(IntegratedLog.txt) .

가

(CreateTable) .



4-3.

```

Function Main()
    Const DTSSQLStgFlag_UseTrustedConnection = 256
    Dim oPKG
    Dim i
    Dim sFolder
    sFolder = "D:\LocalLog\"           //
    Dim fso, f, fl, fc, s
    Set fso = CreateObject("Scripting.FileSystemObject")
    Set f = fso.GetFolder(sFolder)
    Set fc = f.Files
    For Each fl in fc
        Set oPKG = CreateObject("DTS.Package") //DTS Object
        oPKG.LoadFromSQLServer".",,,DTSSQLStgFlag_UseTrustedConnection, , ,
        , "DirImport"                 //
        Set cn = oPKG.Connections("Connection 1") //
        cn.DataSource = sFolder & fl.name
        oPKG.Execute                   //
        oPKG.Uninitialize()
        Set oPKG = Nothing
    Next
    Main = DTSTaskExecResult_Success
End Function

```

4-4.**4.3.3**

(prep.exe)

, 4-2

4-2

				IP	
	778,707	637	192	5,440	35,635
(KB)	167,820	54	8	175	15,236

4.3.4

, , IP ,

OLE DB Provider for SQL Server

(MS SQL

2000)

4-5

4-6

(5Build4DimDW)

(Result.txt),

(IP.txt, Site.txt, Time.txt)

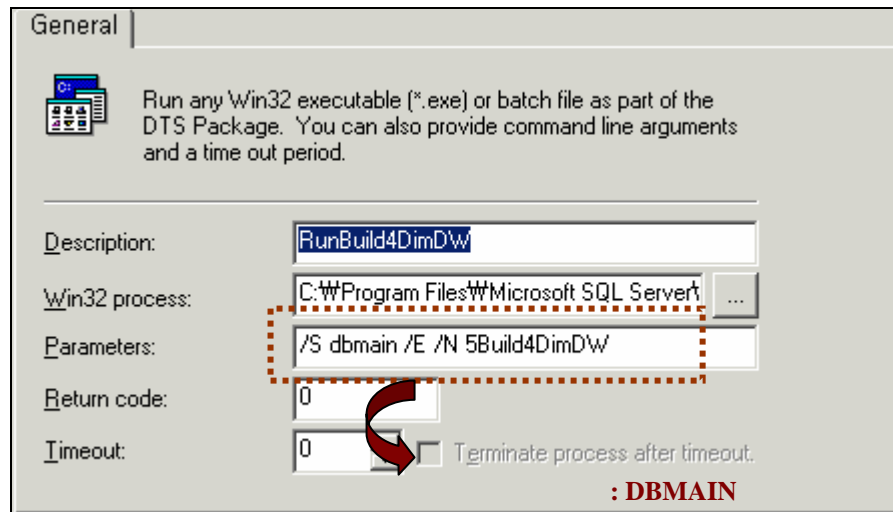
(Customer)

(LogWarehouse)

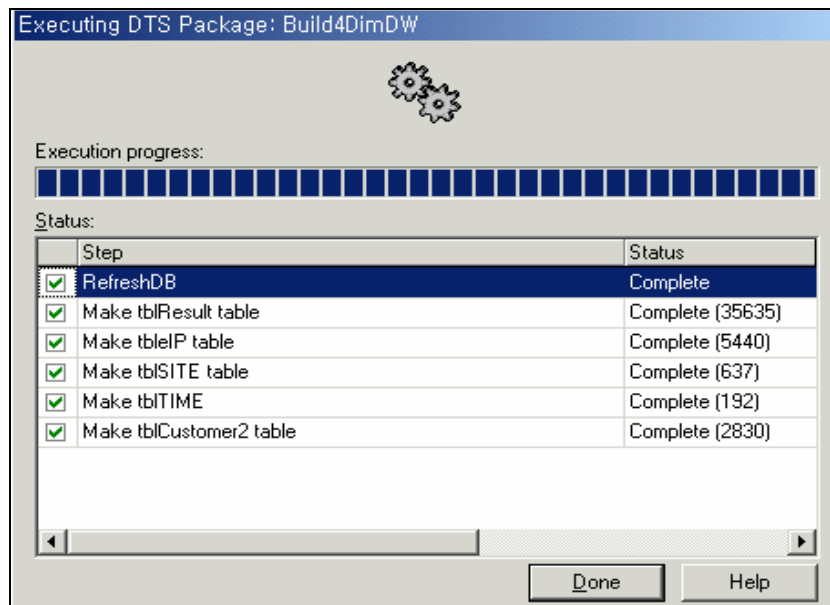
4-7

4-8

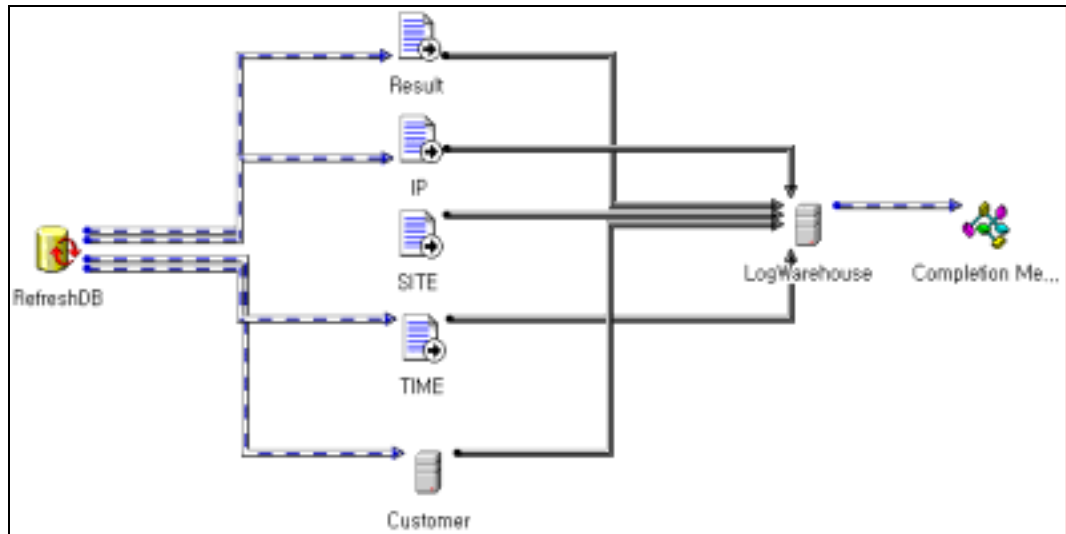
ActiveX



4-5.



4-6.



4-7.

```

C:\Program Files\Microsoft SQL Server\90\Tools\Binn\MSDTL\
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 26000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 26000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 27000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 27000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 28000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 28000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 29000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 29000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 30000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 30000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 31000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 31000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 32000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 32000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 33000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 33000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 34000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 34000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 35000 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 35000
DTSRun OnProgress: DTSStep_DTSDataPumpTask_1; 35635 Rows have been transformed
or copied.; PercentComplete = 0; ProgressCount = 35635
DTSRun OnFinish: DTSStep_DTSDataPumpTask_1
DTSRun OnStart: DTSStep_DTSActiveScriptTask_1
  
```

4-8.

4.3.5 OLAP

OLAP

(Multidimensional OLAP)

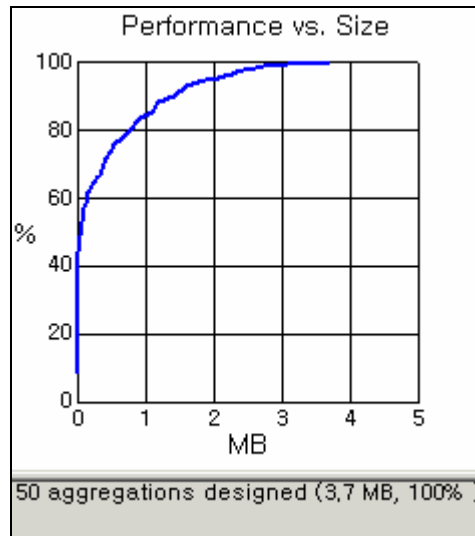
4-9

100%

167.820KB

3.7M

50



4-9.

OLAP

K-

(K-nearest neighbor)

(Microsoft_Clustering)

2000

9

4

OLAP

4.10

MiningModel 'CustomerUsageSegmentation'

```

Execute : CREATE OLAP MINING MODEL [CustomerUsageSegmentation'S]
FROM [LogData] (CASE DIMENSION [Customer] LEVEL [Id]
PROPERTY [Gender] , PROPERTY [Mailing] , PROPERTY [Marital Status] ,
PROPERTY [Yearly Income] , PROPERTY [Num Children At Home] ,
PROPERTY [Education] , PROPERTY [Member Card] , PROPERTY
[Occupation] , PROPERTY [Num Cars Owned] , PROPERTY [Sports Team] ,
MEASURE [Sc Bytes] , MEASURE [Cs Bytes] , MEASURE [Time Taken] ,
MEASURE [Hit Count] , MEASURE [User Count] , MEASURE [Session
Count] )
USING Microsoft_Clustering (CLUSTER_COUNT=4)
Execute : INSERT INTO [CustomerUsageSegmentation'S]

```

4-10. OLAP

OLAP

(CustomerUsageSegmentation)

OLE DB for Data Mining

OPENROWSET

4-11

SELECT FLATTENED

```

[T1].[Customer.Id], [T1].[Customer.Id.Gender], [T1].[Customer.Id.Mailing],
[T1].[Customer.Id.Marital Status], [T1].[Customer.Id.Yearly Income],
[T1].[Customer.Id.Num Children At Home], [T1].[Customer.Id.Education],
[T1].[Customer.Id.Member Card], [T1].[Customer.Id.Occupation],
[T1].[Customer.Id.Num Cars Owned], [T1].[Customer.Id.Sports Team]

```

FROM

[CustomerUsageSegmentation]

PREDICTION JOIN

OPENROWSET(

'MSDataShape',

'Data Provider=MSDASQL.1;Persist Security Info=False;Data

```

Source=LogWarehouse', 'SHAPE {SELECT "Id" AS "Customer.Id", "Gender" AS
"Customer.Id.Gender", "Mailing" AS "Customer.Id.Mailing", "marital_status" AS
"Customer.Id.Marital Status", "yearly_income" AS "Customer.Id.Yearly Income",
"num_children_at_home" AS "Customer.Id.Num Children At Home", "education" AS
"Customer.Id.Education", "member_card" AS "Customer.Id.Member Card", "occupation" AS

```

```

"Customer.Id.Occupation", "num_cars_owned" AS "Customer.Id.Num Cars Owned",
"SportsTeam" AS "Customer.Id.Sports Team" FROM "tblCustomer2" ORDER BY "Id"'
) AS [T1]
ON
[CustomerUsageSegmentation].[Customer.Id] = [T1].[Customer.Id] AND
[CustomerUsageSegmentation].[Customer.Id.Gender] = [T1].[Customer.Id.Gender] AND
[CustomerUsageSegmentation].[Customer.Id.Mailing] = [T1].[Customer.Id.Mailing]
AND[CustomerUsageSegmentation].[Customer.Id.Marital Status]=[T1].[Customer.Id.Marital
Status] AND
[CustomerUsageSegmentation].[Customer.Id.Sports Team] = [T1].[Customer.Id.Sports Team]
    
```

4-11.

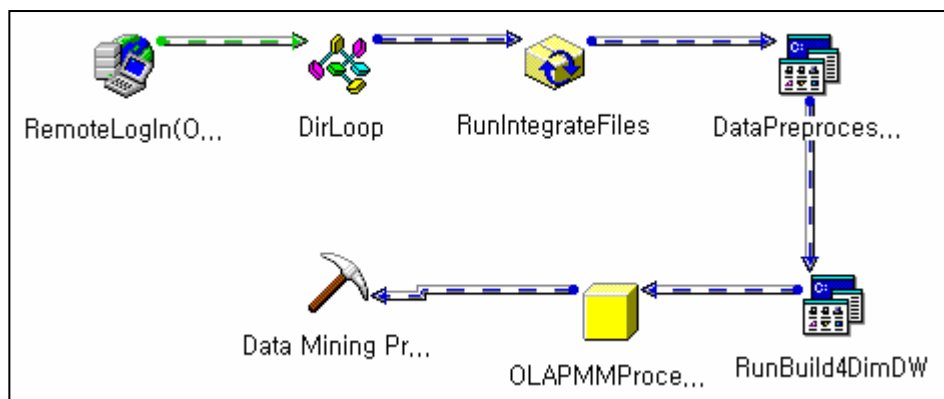
4.3.6

4-12

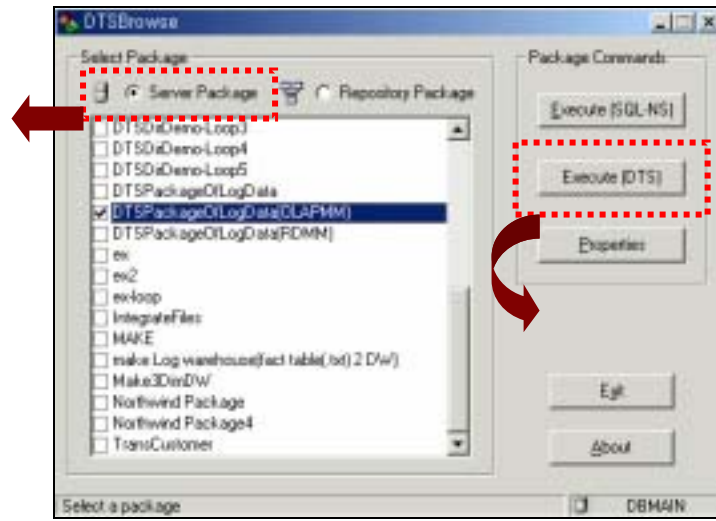
(DTSPackageOfLogDataOLAPMM) . DTS Designer

0

4-13



4-12.



4-13. DTSBrowse : DTS

4.4

OLAP

(Member Card), (Yearly Income) , , ,
IP

4.4.1

Microsoft SQL 2000 Server Analysis Services Manager 가 OLAP

(Drill-Down, Roll-Up, Slice, Dice)

가 .

,

,

가 .

1 :

10

11

가?

(SportsTeam)

가

.

가

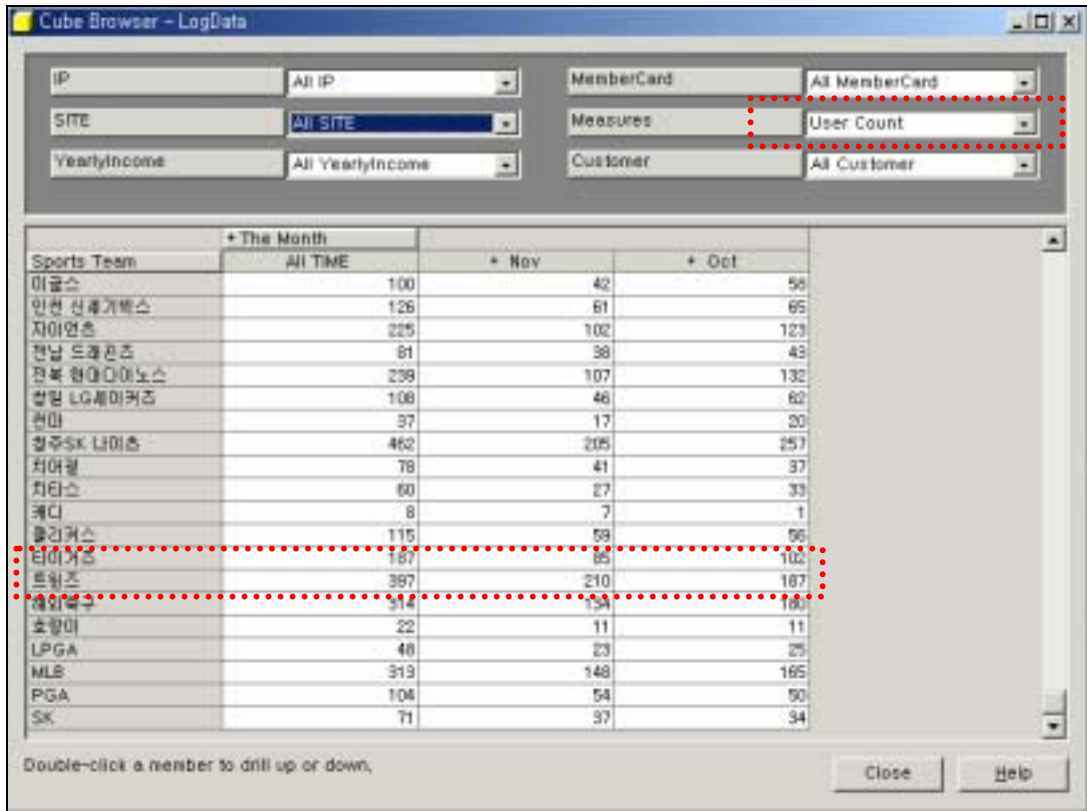
가

.

4-14

1

.



4-14.

1

2 :

가?

가

가

가

. Sports Center, Sports School, Sports Science

가

Sports Tour, Spomed

가

4.15 2

The screenshot shows the 'Cube Browser - LogData' interface. The 'Measures' dropdown is set to 'Session Count'. The pivot table below has the following data:

Category	스포츠 팀	스포츠 팀	스포츠 팀	스포츠 팀
event	9	11	4	5
form	23	24	8	12
GameData_Calendar	1	2		1
home	2		2	1
Index.asp	138	122	79	64
Login_End.asp	1	2	2	1
login_error.asp	2	3	1	1
logout.asp				
MagazineNews	13	11	9	10
main3			1	
Member_Manage		1		1
MySets	1	2	1	1
pop	29	20	15	18
press				
register	8	12	3	9
sitemap			5	
special	1	2		2
specia				
sponed				
SportsCenter	34	21	18	17
SportsSchool	66	72	36	50
SportsScience	89	79	38	48
sportstar	14	22	8	10
SportsTour	13	18	6	4
weather	2	4	2	1

4-15.

2

3 : Beckett

(, ,)

가 가?

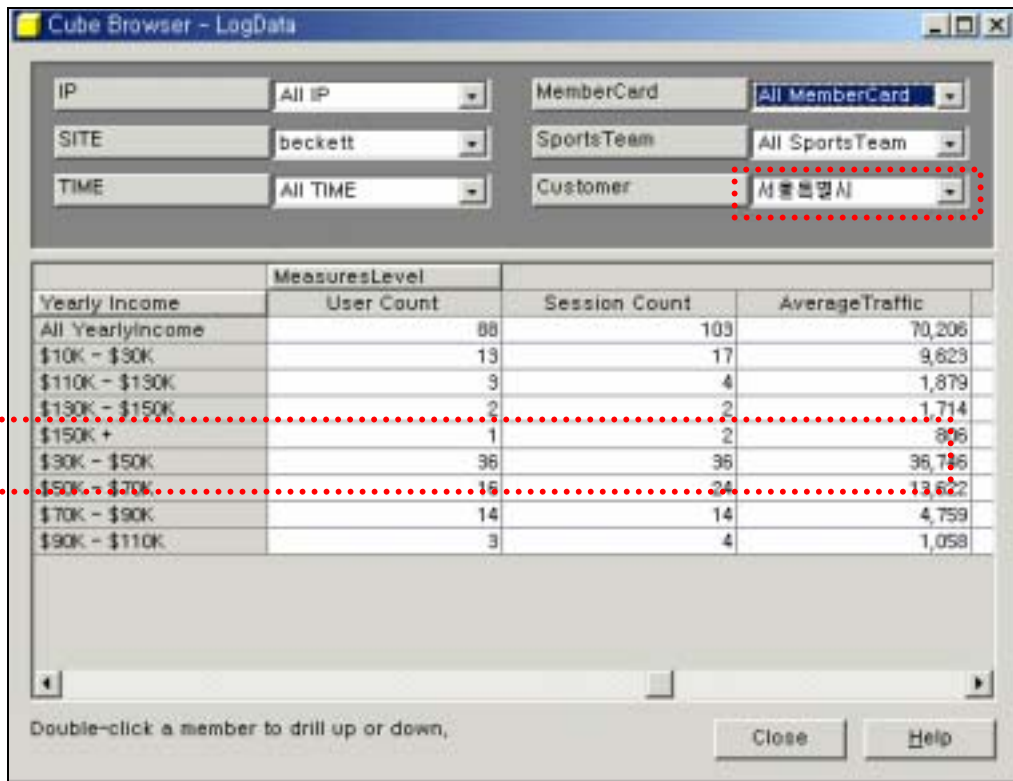
\$30-\$50

가 가

\$150K

가

4-16 3



4-16.

3

4.4.2

2000

10

(iteration) 가

9

4

4.-17

4

(Content

Navigator)

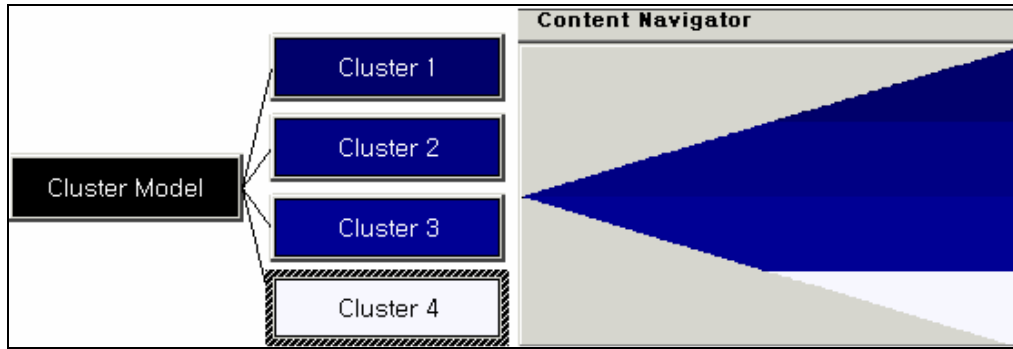
가

100%

1

가 가

4



4-17. K-

4

4-18

4-19

4-18

3

(Node Path)

●

1

1

가 4

\$110-\$130

\$130-\$150, \$150

7

7

Totals Histogram			Totals Histogram		
Value	Cases	Probability	Value	Cases	Probability
(Cluster 1 Total)	1036,30	100,00%	(Cluster 2 Total)	849,33	100,00%
99	149,48	14,42%	99	118,23	13,92%
국가대표	91,06	8,79%	국가대표	74,14	8,73%
라이온스	65,24	6,30%	라이온스	43,99	5,18%
청주SK 나이츠	62,47	6,03%	트윈즈	44,37	5,22%
트윈즈	56,69	5,47%	엔터프라이즈	48,26	5,68%
대전 현대걸리버즈	45,97	4,44%	대전 현대걸리버즈	46,52	5,48%
전북 현대다이노스	34,52	3,33%	청주SK 나이츠	41,88	4,93%
블루윙즈	40,69	3,93%	MLB	41,34	4,87%
오리온즈	37,11	3,58%	해외축구	36,24	4,27%
해외축구	36,44	3,52%	자이언츠	29,67	3,49%
MLB	36,27	3,50%	오리온즈	26,31	3,10%

고객 특성에 따른 사이트 이용도 군집 1 **고객 특성에 따른 사이트 이용도 군집 2**

Totals Histogram			Totals Histogram		
Value	Cases	Probability	Value	Cases	Probability
(Cluster 3 Total)	718,58	100,00%	(Cluster 4 Total)	34,80	100,00%
99	109,30	15,21%	99	5,99	17,22%
국가대표	61,79	8,60%	국내남자	4	11,50%
대전 현대걸리버즈	54,51	7,59%	오리온즈	3	8,62%
라이온스	43,77	6,09%	청주SK 나이츠	3	8,62%
청주SK 나이츠	43,65	6,07%	해외축구	1,80	5,19%
해외축구	31,52	4,39%	자이언츠	2	5,75%
자이언츠	23,71	3,30%	대전 현대걸리버즈	2	5,75%
베어스	28,29	3,94%	국가대표	2	5,75%
엔터프라이즈	24,11	3,36%	PGA	2	5,75%
오리온즈	21,58	3,00%	트윈즈	1	2,87%
트윈즈	25,93	3,61%	이글스	1	2,87%

고객 특성에 따른 사이트 이용도 군집 3 **고객 특성에 따른 사이트 이용도 군집 4**

4-18.

Totals Histogram			Node Path
Value	Cases	Probability	
(Cluster 3 Total)	718,58	100,00%	0 <= Measures,Time Taken <= 11551,8182488064 ,
99	109,30	15,21%	0 <= Measures,Cs Bytes <= 2656,02408674561 ,
국가대표	61,79	8,60%	1 <= Measures,Hit Count <= 6,87743303816885
대전 현대걸리버즈	54,51	7,59%	0 <= Measures,Sc Bytes <= 131344,714955466 ,
라이온스	43,77	6,09%	1 <= Measures,Session Count <= 3,6349955064534 ,
청주SK 나이츠	43,65	6,07%	0 <= Customer,Id,Num Children At Home <= 4,00214798622022 ,
해외축구	31,52	4,39%	1 <= Measures,User Count <= 3,57213300646011 ,
자이언츠	23,71	3,30%	Customer,Id,Marital Status = M ,
베어스	28,29	3,94%	Customer,Id,Yearly Income = \$10K - \$30K ,
엔터프라이즈	24,11	3,36%	Customer,Id,Member Card = Normal ,
오리온즈	21,58	3,00%	Customer,Id,Education = Partial High School ,
트윈즈	25,93	3,61%	Customer,Id,Occupation = Manual ,
			Customer,Id,Member Card = Golden ,
			Customer,Id,Occupation = Skilled Manual ,
			Customer,Id,Sports Team = 대전 현대걸리버즈 ,
			Customer,Id,Sports Team = ,
			Customer,Id,Num Cars Owned <= 0 ,
			Customer,Id,Mailing = 4294967296 ,
			Customer,Id,Sports Team = 이글스 ,
			Customer,Id,Gender = 0

4-19.

- 2
2 가 , ,
, MLB, , 4 ,
\$30-\$50, \$50-\$70 \$70-\$90, 5
, 5 .

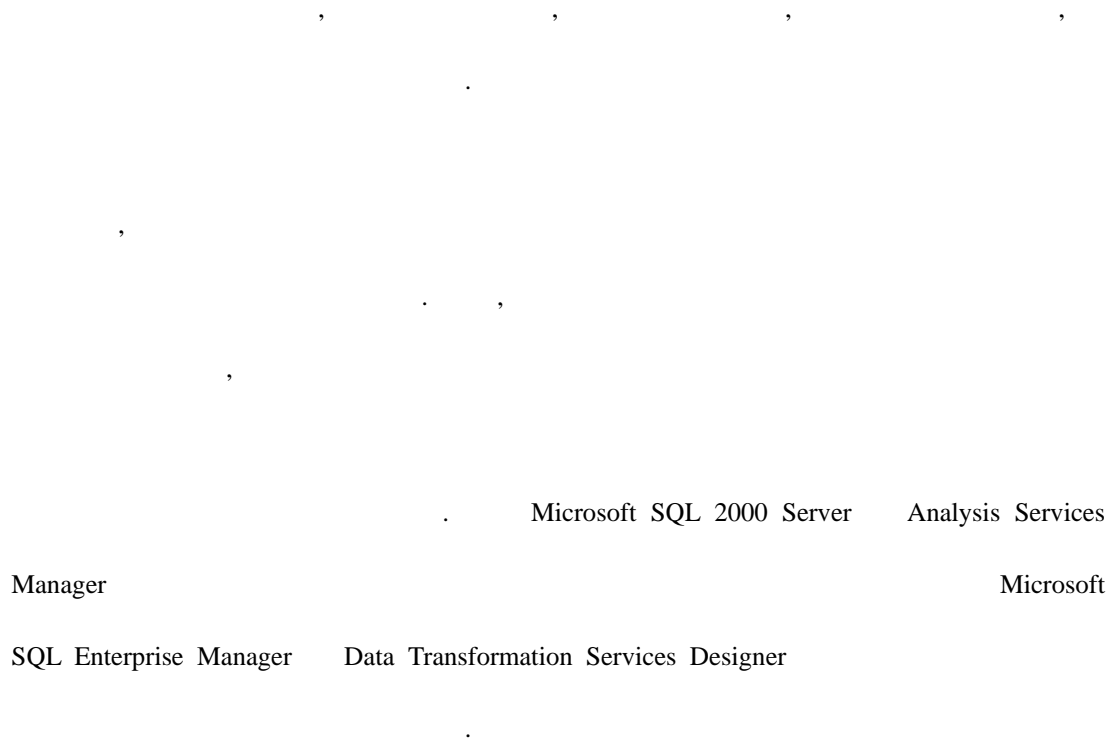
- 3
3 4 , ,
, , ,
\$10-\$30, 4 , 4
.

- 4
4 가 3 , ,
, , , ,
\$70-\$90, 10 , 1
.

가 ()

, ,

V.



- OLAP

OLAP

- 가

가(Refresh)

가

-

IP

IP

-

-

OLAP

COM structure

, Microsoft SQL 2000 Server

가

- OLE DB for Data Mining Specification

. OLE DB for

Data Mining Specification 1.0

가

- [1] Osmar R. Zaiane, Man Xin, Jiawei Han, "Discovering web access patterns and trends by applying OLAP and Data mining technology on web logs ", In Advances in Digital Libraries, pp 19-29, 1998.
- [2] Robert Cooley, Bamshad Mobasher, Jaideep Srivastava, "Data preparation for mining world wide web browsing patterns", the Journal of Knowledge and Information Systems, Vol. 1, No. 1, 1999.
- [3] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, " Web Usage Mining: Discovering and Applications of Usage Patterns from Web Data", ACM SIGKDD, Jan 2000.
- [4] Robert Cooley, Pand-Ning Tan, Jaideep Srivasta, "Discovery of Interesting Usage Patterns from Web Data", Technical Report TR 99-022, University of Minnesota, 1999.
- [5] Bamshad Mobasher, Robert Cooley, Jaideep Srivasta, "Automatic Personalization Based on Web Usage Mining", Technical Report TR99-010, Department of Computer Science, Depaul University, 1999.
- [6] Robert Cooley, Bamshad Mobasher, Jaideep Srivasta, " Web Mining: Information and Pattern Discovery on the World Wide Web", Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), November 1997.
- [7] Raymond Kosala, Hendrik Blockeel, " Web Mining Research: a survey ", ACM SIGKDD 2000, pp 1-15, June 2000. Volume 2, 2000.
- [8] Jian Pei, Jaiwei Han, Behzad Mortazavi-asl, Hua Zhu, " Mining Access Patterns fficiently from Web Logs, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2000.
- [9] OLE DB for Data Mining Specification, version 1.0, Microsoft Corporation, JULY 2000.
- [10] Y. Ma, B. Liu, and C. K. Wong, "Web for Data Mining: Organizing and Interpreting the Discovered Rules using the Web" pp 16-23, Volume 2, SIGKDD 2000.

- [11] M. N. Garafalakis, R. Rastogi, S.Seshadri, K. Shim, “Data mining and the Web: Past, present, and future”, in Workshop on Web Information and Data Management, pp 43-47, 1999.
- [12] <http://www.cs.ualberta.ca/~tszhu/webmining.htm>, Web Mining.
- [13] <http://www.w3.org/pub/WWW/TR/WD-logfile.html>, Extended Log File Format, W3C Working Draft *WD-logfile-960323*.
- [14] <http://www.swynk.com/friends/green/>, Data Transformation Services.
- [15] <http://msdn.microsoft.com/sqlserver/>, MSDN Online Library.
- [16] <http://www.webtrends.com>, WebTrends Corporation.
- [17] <http://www.weblog.com>, Weblog, Inc.
- [18] <http://wum.wiwi.hu-berlin.de/>, Web Utilization Miner.
- [19] <http://www.kdnuggets.com/software/web.html>, Software for Web Mining.
- [20] J. Hipp, U. Guntzer, and G. Nakhaeizadeh, “Algorithms for Association Rule Mining -- A General Survey and Comparison”, SIGKDD 2000.
- [21] <http://www.w3.org/WCA/Activity.html>.

ABSTRACT

Design and Implementation Of Web Usage Pattern Analysis System Using OLAP Mining

*Department of Computer Science & Engineering
Ewha Institute of Science and Technology
Kim Baik Sun*

As a confluence of data mining and WWW technologies, it is possible to perform data mining on web log records collected from the Internet web page access history.

Web usage mining is the application of data mining techniques to discover usage patterns from Web data, in order to understand and better serve the needs of web-based applications

In this paper, we design and implement the web usage pattern analysis system using OLAP Mining. Web usage pattern analysis system consists of three phases, namely data preprocessing, pattern discovery, and pattern analysis and all the process is automated. This paper describes each of these phases in detail.

For data collection, we used user log data saved in IIS web server and customer profile database. For pattern discovery, we used OLAP mining model that using the data source of online analytical processing multidimensional model. For data preprocessing, we presents several data preparation techniques in order to identify unique users and user sessions and above all, implemented dimensional attribute automatic setting algorithm for multidimensional data model.

Web usage pattern analysis system using OLAP Mining composed of log file transport package, integrated log file generation package, data preprocessing execution package, data warehouse building

package and prediction query execution package. We described the results of pattern analysis about example web site.